DOCUMENT RESUME

ED 295 959 TM 011 733

AUTHOR Chambers, Donald B.; Holzemer, William L.
TITLE Validity and Reliability of the Graduate Program Self-Assessment (GPSA) Instruments for Evaluating Nursing Doctoral Education.
SPONS AGENCY Public Health Service (DHHS), Rockville, Md. Div. of Nursing.
PUB DATE Apr 88
GRANT 1-RO-NU00967
NOTE 92p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Alumni; *Construct Validity; *Content Validity; *Doctoral Programs; Educational Quality; Graduate School Faculty; Graduate Students; Higher Education; *Nursing Education; Program Evaluation; *Psychometrics; *Test Reliability
IDENTIFIERS *Graduate Program Self Assessment Questionnaires

ABSTRACT

This paper presents the investigation of the psychometric properties of the Graduate Program Self-Assessment (GPSA) questionnaires, developed by the Educational Testing Service, for the evaluation of nursing doctoral education. The data are from the 1984 cooperative program evaluation from 326 faculty, 659 doctoral students, and 296 alumni. The primary emphasis was on assessing content validity, factorial (construct) validity, concurrent (criterion-related) validity, and internal-consistency reliability (coefficient alpha) of the 16 summary scales for the faculty, student, and alumni questionnaires. In general, the questionnaires demonstrated satisfactory validity and reliability. The analysis provided supportive evidence that there are indeed multiple dimensions of quality in doctoral education and that those dimensions can be measured with the GPSA questionnaires and demonstrated to correlate with other measures of quality. Specific recommendations for summary scale changes were made to improve the psychometric properties of the scales. Fifteen tables are provided. (Author/TJH)

Validity and Reliability of the Graduate Program Self-Assessment (GPSA) Instruments for Evaluating Nursing Doctoral Education

Donald B. Chambers

William L. Holzemer

School of Nursing

University of California, San Francisco

This project was supported by the grant, "Quality Indicators of Nursing Doctoral Programs," DHHS, PHS, Division of Nursing, 1 RO NU00967.

Paper presented at the 1988 Annual Meeting of the American Educational Research Association, New Orleans, April 5-9.

Running head: GPSA VALIDITY AND RELIABILITY

Abstract

This paper presents the investigation of the psychometric properties of the Graduate Program Self-Assessment (GPSA) instruments for evaluation of nursing doctoral education, based on the data collected in the 1984 cooperative program evaluation from 326 faculty, 659 doctoral students, and 296 alumni. The primary emphasis was on assessing content validity, factorial (construct) validity, concurrent (criterion-related) validity, and internal-consistency reliability (coefficient alpha) of the 16 summary scales for the faculty, student, and alumni questionnaires. In general, the questionnaires demonstrated satisfactory validity and reliability. The analyses provided supportive evidence that there are indeed multiple dimensions of quality in doctoral education, and that those dimensions can be measured with the GPSA questionnaires and demonstrated to correlate with other measures of quality. Specific recommendations for summary scale changes were made to improve the psychometric properties of the scales.

Validity and Reliability of the Graduate Program Self-Assessment (GPSA) Instruments for Evaluating Nursing Doctoral Education

This paper presents the investigation of the psychometric properties of the Graduate Program Self-Assessment (GPSA) instruments for evaluation of nursing doctoral education. The primary emphasis was on assessing content validity, factorial (construct) validity, concurrent (criterion-related) validity, and internal-consistency reliability (coefficient alpha) of the 16 summary scales for the faculty, student, and alumni questionnaires.

Evaluation of program quality has been an issue in graduate education in the United States. Traditionally, professional reputation among experts, such as deans, has been used to estimate quality (Blau & Margulies, 1974; Chamings, 1984; Holzemer, 1982). Clark (1983) stated, "Though carefully done and useful in a number of ways, these ratings have been critized for their failure to reflect the complexity of graduate programs, their tendency to emphasize the traditional values that are highly related to program size and wealth, and their lack of timeliness or currency" (p. 1). More recently, use of multiple indicators has developed (Gourman, 1980). One set of parameters for determining the dimensions of quality of doctoral programs has been reported by

Clark, Hartnett, and Baird (1976) in their study conducted by the Educational Testing Service (ETS).

The ETS project was developed in response to the problems associated with traditional ratings of excellence of graduate programs by deans. It was hypothesized that there were multiple dimensions of quality and that those dimensions could be measured and demonstrated to correlate with other measures of quality. The Clark, Hartnett, and Baird (1976) study documented the development and implementation of scales to measure dimensions of quality in doctoral education. In particular, questionnaires were designed for surveying faculty members who taught doctoral students, enrolled doctoral students, and recent doctoral program graduates. Each questionnaire asked respondents to rate a variety of program characteristics based on their experiences or observations in the department, and to provide information about their own activities, achievements, and backgrounds. Pelczar (1985) stated, "The new underlying assumption is that the perceptions and judgments of faculty, students, and alumni can contribute to a better understanding and quality of a department or program" (p. 98).

The questionnaires were used to collect information about the doctoral programs of chemistry, history, and psychology at 25 diverse American universities. These three disciplines were selected for study because they represented major areas of academic endeavor, had large and well-established doctoral

programs, and were different enough to provide a practical test of whether it was feasible to use one set of data collection instruments in the assessment of doctoral programs in several fields. The results of the study indicated that common questionnaires could be used to obtain dependable and useful information about many important program characteristics in different disciplines, but that there were enough differences among the fields to recommend discipline-specific comparison data for several of the variables. Evidence was provided for the reliability of averaged responses to individual questionnaire items, the reliability of composite scores used to summarize judgments about program functioning in a number of areas, and the validity of survey results as indicators of doctoral program quality (Clark, 1983; Clark, Hartnett, & Baird, 1976). Specific information on the validity, reliability, and other psychometric aspects of the GPSA questionnaires are provided in the Instruments section of the paper.

The research questionnaires from the ETS project were adapted for use by the ETS Graduate Program Self-Assessment (GI A) Service. Questionnaires for faculty, students, and alumni were designed to obtain information about important quality-related program characteristics in seven areas: program purposes, faculty training and accomplishments, student ability and performance, resources, academic and social environments of the program,

program processes and procedures, and alumni achievements. Judgments about individual items are combined to form 16 summary scale scores. Where appropriate, identical items appear on all three questionnaires, thus allowing programs to compare the opinions of faculty, students, and alumni (Clark, 1983).

In 1979, 18 of the 22 nursing doctoral programs then in existence participated in a cooperative program evaluation (Barhyte & Holzemer, 1981; Holzemer, 1978; Holzemer & Barhyte, 1979; Holzemer, Barhyte, & Clark, 1980). The primary evaluation tools were the ETS GPSA questionnaires. Results reported by ETS included confidential reports to each participating program. Programs also received group comparative data compiled from all participating programs, and comparative data from the ETS study of chemistry, history, and psychology doctoral programs. The cooperative program evaluation found variation among nursing doctoral programs, but comparison of nursing faculty and student perceptions with those of faculty and students in chemistry, history, and psychology revealed more similarity than differences between nursing and the other disciplines. A major limitation of the study, however, was the fact that many of the participating programs had been recently established and had few students and/or alumni. Furthermore, only group (program-level) summary data was available from ETS, which limited investigations into the validity

and reliability of the GPSA questionnaires for evaluating doctoral education in nursing.

In 1984, 25 of the 29 nursing doctoral programs then in existence participated in a follow-up study of the 1979 cooperative program evaluation (Holzemer, 1987). The primary evaluation tools were, again, the ETS GPSA questionnaires. Special arrangements were made with ETS for the 1984 study to provide anonymous, individual respondent data so that appropriate statistical methods could be employed to investigate validity and reliability of the GPSA scales. Normally, the ETS GPSA Service produces only program-level reports summarizing group responses.

The psychometric properties of the GPSA instruments were investigated for use in evaluation of nursing doctoral education, based on the data collected in the 1984 study from 326 faculty, 659 doctoral students, and 296 alumni. The primary emphasis was on assessing content validity, factorial (construct) validity, concurrent (criterion-related) validity, and internal-consistency reliability (coefficient alpha) of the 16 summary scales for the faculty, student, and alumni questionnaires.

Method

Sample

All doctoral programs in nursing were invited to participate in the study during the summer of 1983. There were 29 eligible

programs; an eligible program was defined as one that was commited to admitting students to the program by fall, 1985. Twenty-five (86%) of the programs agreed to participate in the study. No reason was requested from the four non-participating programs.

The overall response rates for programs and individuals are presented in Table 1. The number of usable questionnaires returned was 326 for faculty (55% response rate), 659 for students (54% response rate), and 296 for alumni (60% response rate). A usable questionnaire is defined by ETS as any GPSA questionnaire having valid responses to 10 or more questions across Parts I and II combined.

Insert Table 1 about here

Instruments

The Graduate Program Self-Assessment (GPSA) questionnaires developed by Educational Testing Service were used. As stated in the Introduction to the paper, the GPSA questionnaires are adaptions of instruments used in the mid-1970s tc study the dimensions of quality in doctoral education. Developed in cooperation with committees of graduate deans and faculty members, the questionnaires were designed to obtain information about important quality-related program characteristics in seven areas: program purposes, faculty training and accomplishments, student

ability and performance, resources, academic and social environments of the program, program processes and procedures, and alumni achievements.

The core of each questionnaire consists of approximately 60 statements concerning characteristics of the program, generally with agree-to-disagree or poor-to-excellent ratings as response options. Judgements about individual items are combined to form 16 summary scale scores to describe several areas of program functioning. Summary scales 1 through 14 are reported as averages of the item responses making up those scales. Summary scales 15 and 16 are reported in percentages rather than mean scores; these percentages represent the number of items to which faculty responded positively in the list of individual research and professional activities presented. Respondents must complete a minimum number of items in a scale to receive a summary scale score. Descriptions of these summary scales and the number of individual items included in each scale are contained in Table 2.

Insert Table 2 about here

Evidence concerning the psychometric reliability and validity of the GPSA instruments is based on the use of similar, experimental questionnaires in the assessment of seventy-three doctoral programs in the fields of chemistry, history, and

psychology (Clark, Hartnett, & Baird, 1976) and is summarized by Clark (1983) in the GPSA Handbook for Users. The median reliability (intraclass correlation) for the summary scales was .76, with a range from .46 to .90. Tests of scale homogeneity or internal consistency (coefficient alpha) ranged from .68 to .93, with a median of .83.

Intercorrelations of department scores on the summary scales were generally positive and moderate, with a median correlation coefficient of .31. In general, student summary scale scores were more highly intercorrelated than those of faculty and alumni. Clark (1983) stated:

> Clearly, students who had a high opinion of their doctoral program in one of these areas tended to respond favorably in the other areas as well. However, none of the correlations were sufficiently high to preclude the possibility of within-program differences in scale scores, and the areas of program functioning were considered sufficiently distinct conceptually to warrant separate assessment. It was felt that, as instruments for program review and improvement, separate scores on overlapping indicators, such as Quality of Teaching and Faculty Concern for Students, would be more useful than scores on a smaller number of scales selected primarily for their psychometric independence. (p. 13)

Content and concurrent validity of the GPSA instruments was examined in a number of areas and is summarized in the technical report of the research (Clark, Hartnett, & Baird, 1976). Research evidence indicated that responses to GPSA questionnaire scales should be valid and useful indicators of program status.

In addition to the individual items comprising the 16 summary scales, the GPSA instruments include questions about the respondent's activities and background characteristics, such as faculty scholarly and professional productivity, student educational experiences and career interests, and alumni employment and professional accomplishments. Additional items for faculty, students, and alumni were developed by a national nursing advisory group for the evaluation study. These items, judged as unique and important to nursing doctoral education but not directly addressed in the GPSA instruments, were included in separate questionnaires and mailed with the GPSA questionnaires.

Procedure

Questionnaires were mailed during the winter of 1984 to faculty and students of the participating programs. Alumni questionnaires were mailed approximately one month later to avoid a faculty member simultaneously receiving both the faculty and alumni questionnaire.

ETS tabulated all questionnaires and provided anonymous, ... al faculty, student, and alumni respondent data, as well

as the standard program-level, group data. The validity and reliability analyses were performed utilizing either individual respondent or program-level data. Aggregated, program-level data were used when external measures, such as reputational ratings, were based on the program as the unit of analysis. The level of data used and the related sample size are reported with each analysis in the Results section.

Results and Discussion

Results of the investigation into the psychometric properties of the 16 summary scales of the Graduate Program Self-Assessment (GPSA) questionnaires are presented in four parts: content validity, factorial construct validity, internal-consistency reliability, and concurrent validity. Where appropriate, results are reported separately for the faculty, student, and alumni questionnaires.

Content Validity

As part of the 1979 cooperative program evaluation (Holzemer, 1978; Holzemer & Barhyte, 1981), a thorough review of the GPSA questionnaires was performed by three nursing experts directing doctoral education programs in nursing. Doctoral nursing programs polled prior to the 1979 evaluation study were shown complete sets of the questionnaires before being asked for a commitment for participation in the project. In addition, national advisory

committees to both the 1979 and 1984 evaluation studies carefully reviewed the GPSA questionnaires. In general, all experts found the GPSA instruments to have content appropriate and valid for evaluating a variety of dimensions of quality common to all doctoral programs. The experts, however, noted that the questionnaires have several limitations. First, the questionnaires assess only perceptions of quality. Second, questionnaire items fail to assess areas of concern to a practice discipline, such as advanced clinical practice in nursing. Third, the questionnaires do not assess the goal of nursing doctoral education, that is, to increase the scientific body of knowledge within nursing.

The GPSA service allows up to 10 locally developed, fixed-format items to be added to each of the questionnaires. This option provides programs the opportunity to further increase the content validity of the GPSA questionnaires by adding items that are of interest or significance at the local, state, or national level. This option is of particular importance to practice professions, such as nursing, for it enables clinical aspects of the profession to be assessed in the GPSA questionnaires. The national advisory committees to both the 1979 and 1984 evaluation studies formulated additional questions judged unique and important to nursing doctoral education. The addition of program-specific items did not affect the reliability of the

GPSA, because responses to the optional items were not included with any of the summary scale scores.

Factorial Construct Validity

Factorial construct validity of the GPSA summary scales was investigated for each questionnaire (faculty, student, and alumni) at two levels. At the item level, separate factor analyses were performed within each of the 16 summary scales. The primary purpose of these within-scale, item-level analyses was to determine the factorial complexity of the separate summary scales, that is, the degree of scale homogeneity/heterogeneity. It was anticipated that these analyses would support and possibly add to the results of the internal-consistency reliability analyses, discussed in the next section of the paper. At the summary scale level, second-order factor analyses (Allen & Yen, 1979) were performed using the 16 summary scale scores. The primary purpose of these scale-level analyses was to investigate the convergent/discriminant construct validity of the 16 summary scales as measures of the hypothesized multidimensional concept of quality in doctoral education.

Faculty questionnaire. Descriptive statistics and intercorrelations for the GPSA summary scales for faculty are reported in Table 3, based on the 299 faculty who had ETS-calculated scale scores for all 11 faculty scales; dashed lines indicate scales not applicable to the faculty questionnaire.

contained one or two items each with loadings considerably less than those of the other items in the scale, indicative of weaker interrelationships, and consideration should be given to dropping these items from the scales; they included items I-3 and I-7 in Scale 1 (loading .33 and .31, respectively), I-9 in Scale 6 (loading .38), and I-12 in Scale 11 (loading .42).

Scale 12, which demonstrated one factor, had one item (I-5) with a low loading (.27), particularly when compared to loadings ranging from .61 to .84 for the other five items in the scale. The intercorrelations of item I-5 with the other items in scale were low, ranging from .15 to .24, indicating that perhaps I-5 should be dropped from Scale 12. Although Scale 16 demonstrated one factor, with the exception of item III-5 (loading .60), item loadings were relatively low, ranging from .26 to .44. The intercorrelations of all five items making up the scale were very low, ranging from .06 to .28, indicating a strong degree of item heterogeneity. This finding was also supported by the results of the internal-consistency reliability analysis, discussed in the next section of the paper.

With two retained factors after the initial extraction, only Scale 15 (Faculty Research Activities) demonstrated factorial complexity with a moderately strong first factor, defined by items III-9, III-10, and III-11, and a somewhat weaker second factor, defined by items III-3, III-7, and III-8. The three items loading

contained one or two items each with loadings considerably less than those of the other items in the scale, indicative of weaker interrelationships, and consideration should be given to dropping these items from the scales; they included items I-3 and I-7 in Scale 1 (loading .33 and .31, respectively), I-9 in Scale 6 (loading .38), and I-12 in Scale 11 (loading .42).

Scale 12, which demonstrated one factor, had one item (I-5) with a low loading (.27), particularly when compared to loadings ranging from .61 to .84 for the other five items in the scale. The intercorrelations of item I-5 with the other items in scale were low, ranging from .15 to .24, indicating that perhaps I-5 should be dropped from Scale 12. Although Scale 16 demonstrated one factor, with the exception of item III-5 (loading .60), item loadings were relatively low, ranging from .26 to .44. The intercorrelations of all five items making up the scale were very low, ranging from .06 to .28, indicating a strong degree of item heterogeneity. This finding was also supported by the results of the internal-consistency reliability analysis, discussed in the next section of the paper.

With two retained factors after the initial extraction, only Scale 15 (Faculty Research Activities) demonstrated factorial complexity with a moderately strong first factor, defined by items III-9, III-10, and III-11, and a somewhat weaker second factor, defined by items III-3, III-7, and III-8. The three items loading

on the first factor assess grant support of faculty research, while the three items loading on the second factor relate to recognition of excellence in research and scholarly writing.

A second-order principal axis factor analysis using the 11 faculty summary scales (n=299) extracted two factors (eigenvalues 5.95 and 1.28, explained variance 54% and 12%, respectively). Both before and after varimax rotation, the first nine scales (1, 2, 4, 5, 6, 7, 8, 11, and 12) clearly defined the first factor, with rotated item loadings ranging from .52 to .87; Scales 15 and 16 clearly defined the second factor, with item loadings of .57 and .46, respectively.

These findings are not unexpected, based on the intercorrelations of the faculty summary scale scores presented in Table 3. Intercorrelations for Scales 1 through 12 ranged from .35 to .81, with a median of .62. Scales 15 and 16 correlated moderately (.30) with each other, but only weakly with the other nine faculty scales (.03 to .22). Apparently, there is some divergent (discriminant) construct validity in the 11 faculty summary scales, with the first 9 measuring various aspects of the academic program environment and the last 2 measuring faculty productivity. These findings provided empirical support to the division of the faculty GPSA summary scales into sets of environmental and productivity variables, as investigated in recent papers by Holzemer and Chambers (1986, 1988).

Student questionnaire. Descriptive statistics and intercorrelations for the GPSA summary scales for students are reported in Table 5. Results for Scales 1 through 9 were based on the 538 students who had ETS-calculated scale scores for all 9 of these student scales; results for Scale 10 were based on the 252 students who had been a research or teaching assistant in their department and had ETS-calculated scale scores for all 10 of the student scales. Results of the principal axis factor analysis, with varimax rotation, are summarized in the right half of Table 6, based on 293 students for Scales 1 through 9 and 281 students for Scale 10.

Insert Tables 5 and 6 about here

Only one primary factor was extracted for all 10 of the student summary scales. For Scales 2 through 10, all item loadings were consistently greater than .30, indicating the likelihood of within-scale homogeneity of all items making up those scales. Only Scale 1 contained items (I-3 and I-7) with loadings (.33 and .14, respectively) considerably less than those of the other items in the scale (loadings .63 to .80), and consideration should be given to dropping them from the scale. These same two items were recommended for deletion in Scale 1 of the faculty questionnaire.

A second-order principal axis factor analysis using the 10 student summary scales (n=252) also extracted only one primary factor (eigenvalue 5.93, explained variance 59%), with item loadings ranging from .40 to .88. Scale intercorrelations ranged from .20 to .81, with a median of .55. Although Scales 7, 8, and 10 demonstrated somewhat weaker interrelationships among themselves and with the other student summary scales, the findings of the second-order factor analysis tended to confirm the existence of convergent construct validity of the 10 student scales as separate, though sufficiently related measures of the overall concept of quality in doctoral education.

Alumni questionnaire. Descriptive statistics and intercorrelations for the GPSA summary scales for alumni are reported in Table 7, based on the 260 alumni who had ETS-calculated scale scores for all 10 alumni scales. Results of the principal axis factor analysis, with varimax rotation, are summarized in the right half of Table 8. Analyses for Scales 1 through 13 were based on the 207 alumni who answered all 54 items comprising Scales 1 through 13 of the alumni questionnaire; because of scoring rules for Scale 14 set by ETS, analyses for Scale 14 were based on the 68 alumni who had been a research or teaching assistant in their department and answered all 13 items comprising Scale 14.

Insert Tables 7 and 8 about here

Only one primary factor was extracted for 9 of the 10 alumni summary scales. No solution was reached for Scale 9, but intercorrelations of the three items comprising this scale ranged from .32 to .64, indicating moderate homogeneity of these scale items. With the exception of Scale 1, all item loadings were consistently greater than .30, again indicating the likelihood of within-scale homogeneity of all items making up those scales. Scale 1 contained two items (I-3 and I-7) with loadings (.37 and .24, respectively) considerably less than those of the other items in the scale (loadings .63 to .77). Once again, consideration should be given to dropping them from the scale, for these same two items were recommended for deletion in both the faculty and student questionnaires.

Because of the small sample size (n=68) used for the factor analysis of Scale 14, the factor analysis was rerun using only alumni who answered the first 11 items comprising Scale 14, that is, the alumni who had not been a research or teaching assistant in their department and, therefore, did not answer items IV-12 or IV-13. Based on 256 alumni, the results were very similar to those based on the smaller sample of 68. Once again, only one

primary factor was extracted, with item loadings ranging from .36 to .75.

A second-order principal axis factor analysis using the 10 alumni summary scales (n=260) also extracted only one primary factor (eigenvalue 6.19, explained variance 62%), with item loadings ranging from .41 to .90. Scale intercorrelations ranged from .20 to .80, with a median of .60. As with the student questionnaire, the findings of the second-order factor analysis tended to confirm the existence of convergent construct validity of the 10 alumni scales as separate, though sufficiently related measures of the overall concept of quality in doctoral education.

Internal-Consistency Reliability

Results of the internal-consistency reliability analyses for the GPSA summary scales are summarized separately for faculty, students, and alumni in the left half of Tables 4, 6, and 8, respectively. The summary data reported for each scale include the minimum, maximum, and mean interitem correlation for that scale, plus coefficient alpha.

Other things being equal, the more reliable a measuring procedure is, the better. It is difficult, however, to specify a single level of reliability that should apply in all situations. Discussions by Carmines and Zeller (1979), Polit and Hungler (1978), and Thorndike and Hagen (1977) supported the notion of higher reliability coefficients (.80 to .90, or higher) as being

necessary for instruments used for making decisions about individuals, and lower coefficients (.60 to .70) as being sufficient for decisions involving group-level data. Because ETS reports only group-level results to programs using the GPSA questionnaires, and it is generally true that aggregated variables are much more reliable than would be the case with individual measurements, it was decided to consider as acceptable all GPSA summary scales with coefficient alphas .60 or greater.

Faculty questionnaire. The reliability analyses were based on the 236 faculty who answered all 60 summary scale items of the faculty questionnaire. Ten of the 11 faculty summary scales had coefficient alphas greater than .60, demonstrating satisfactory levels of internal-consistency reliability. Only Scale 16 demonstrated a lack of internal consistency (alpha .49). This was not surprising given the strong degree of item heterogeneity, as indicated by the low intercorrelations (.06 to .28) of all five items making up the scale.

By taking into account the findings of both the within-scale, item-level factor analyses and the internal-consistency reliability analyses, four of the faculty summary scales could have their coefficient alphas increased by dropping one or two items that did not seem to relate to the other items within the scale. These included Scale 1 (drop I-3 and I-7, new alpha .78), Scale 6 (drop I-9, new alpha .87), Scale 11 (drop I-12, new alpha

.83), and Scale 12 (drop I-5, new alpha .82). Consideration should be given to splitting Scale 15 into two new scales of three items each, and then adding new items tapping the concepts of the two new scales: grant support of faculty research and recognition of faculty excellence in research and scholarly writing. Finally, the reliability of Scale 7, which has only three items, could be improved by the addition of more items tapping the same concept of available resources.

Student questionnaire. The reliability analyses were based on 293 students for Scales 1 through 9 and 281 students for Scale 10. All ten of the student summary scales had coefficient alphas greater than .60, demonstrating satisfactory levels of internal-consistency reliability for group-level data. Two of the student summary scales could have their coefficient alphas increased by dropping one or two items. These included Scale 1 (drop I-3 and I-7, new alpha .76) and Scale 8 (drop I-11, new alpha .70). The reliability of Scale 7, which has only two items, could be improved by the addition of more items tapping the same concept of available resources.

Alumni questionnaire. The reliability analyses were based on 207 alumni for Scales 1 through 13 and 68 alumni for Scale 14. Nine of the ten alumni summary scales had coefficient alphas greater than .60, demonstrating satisfactory levels of internal-consistency reliability for group-level data. It is not

surprising that Scale 7, with only two items, had an alpha of .57. As with the faculty and student questionnaires, the reliability could be improved by the addition of more items tapping the same concept of available resources. Finally, the coefficient alpha of Scale 1 could be increased by dropping I-3 and I-7 (new alpha .75).

Concurrent Validity

Concurrent validity of the GPSA summary scales was investigated for each questionnaire (faculty, student, and alumni) by correlating the scale scores with various "internal" and "external" criterion measures. Internal measures included responses to selected items within the ETS GPSA questionnaires that were not included in the 16 summary scales, plus selected items from those developed by the national advisory committee for the evaluation study. For ease of presentation and interpretation of the results, the selected items were divided into four general categories: academic and social environment, resources and management, scholarship and productivity, and faculty ranking of doctoral programs. Correlations involving the first three internal sets of items were performed using the individual respondent as the unit of analysis; correlations involving the faculty ranking of doctoral programs in nursing were performed using the program as the unit of analysis and are presented with the external measures.

External criterion measures included Chamings' rankings of nursing schools by 252 deans and other nursing academics and professionals (Chamings, 1984), Grout's tabulation of the number of faculty publications in scholarly nursing journals from 1978 to 1982 (Grout, 1985; Grout, personal communication, March, 1986), and the number of Division of Nursing, DHHS, funded research grants from 1979 to 1983 (Bloch, personal communication, 1985). All correlations involving the external measures were performed using the program as the unit of analysis. Descriptive statistics and Spearman rank-order intercorrelations for the three external criterion measures and the faculty ranking of doctoral programs (internal measure) are reported in Table 9. To eliminate negative correlations with variables not based on rankings, the two ranking variables were recoded so that high rankings were associated with a high number (e.g., 25) rather than the traditional low number (e.g., 1). The intercorrelations among the four criterion measures were statistically significant at the .05 level of significance (two-tailed), and demonstrated moderate to high levels of interrelationships.

Insert Table 9 about here

The ranking of nursing schools by Chamings was an update of an earlier survey by Blau and Margulies (1974) and was based on a

1982 survey of all accredited nursing schools in the United States. This differed from the faculty rankings in the 1984 cooperative program evaluation, which were based only on doctoral nursing programs. These two rankings, however, were highly correlated (Spearman rank-order correlation = .84, $p <= .001$) for the 22 programs ranked by both groups.

Grout's tabulation of faculty publications was selected over Hayter's (1984) tabulation, because Grout used only the 3 nursing journals (Nursing Research, Research in Nursing and Health, and Western Journal of Nursing Research) rated highest in scholarship by deans of nursing schools (Fagin, 1982). Hayter used 13 nursing journals intended for a general nursing audience. Grout (1985) commented, "Only 7 of the 13 journals selected by Hayter . . . were recognized by deans of nursing schools as rating 'highest in overall quality,' and none of the 9 that accounted for 83% of the articles tabulated were rated 'highest in scholarship' [by Fagin]" (p. 204). Although not reported in Table 9, the tabulations by Grout and Hayter were highly correlated (Spearman rank-order correlation = .76, $p <= .001$) for the 20 programs ranked by both studies and included in the 1984 cooperative program evaluation.

Faculty questionnaire. Descriptive statistics and correlations of criterion measures with GPSA summary scale scores for faculty are reported in Tables 10 and 11. Table 10 includes internal criterion measures based on the individual respondent as

the unit of analysis; Table 11 includes faculty ranking of doctoral programs (internal measure) and the three external criterion measures, which are based on the program as the unit of analysis. Only correlations with an absolute value of .30 or greater are reported in the tables, based on recommendations of Cohen (1977), that $\underline{r}$ = .30 represents a medium effect size for "real-world " significance, and of Guilford (1965), that $\underline{r}$ = .30 is typical of criterion validity coefficients for psychological tests. Because of the large sample sizes for individual respondent data, all Pearson product-moment correlations of .30 or greater were significant at $\underline{p}$ <= .001, two-tailed. Program-level data were based on much smaller samples (maximum 25), but considered more reliable because they were comprised of aggregated data rather than individual measurements. Therefore, when based on program-level data, even non-significant Spearman rank-order correlations of .30 or greater are reported, and correlations significant at $\underline{p}$ <= .05, two-tailed, are underlined.

Insert Tables 10 and 11 about here

With the individual faculty member as the unit of analysis (see Table 10), associations between internal criterion measures and the GPSA summary scales demonstrated moderate evidence for concurrent validity of the scales. Within the set of academic and

social environment measures, higher academic rank and tenure were related to higher reported levels of faculty professional activities (Scale 16). Percent time spent on research and scholarly work was positively related to reported levels of faculty research activities (Scale 15). Within the set of items developed by the national advisory committee, faculty were asked to rate the degree to which five descriptors were characteristic of the environment of their doctoral program; three of these descriptors ("scholarly," "healthy," and "prestigious") were positively correlated with nearly all of the faculty ratings of their program based on the GPSA academic program environment scales (Scales 1 through 12). Finally, faculty perception of the degree to which their colleagues are involved in an active program of research was positively related to all but one of the GPSA environmental scales. This is an important indicator of the quality of the academic program environment at the doctoral education level.

Criterion measures of program resources and management were limited. Within the set of advisory committee items, faculty were asked to indicate the availability and rate the adequacy of six support services in their setting. Ratings of secretarial support, travel monies, and release time for scholarly activity were positively correlated with ratings of the faculty work environment (Scale 12). Rating of release time was also

positively related to ratings of the curriculum (Scale 5) and departmental direction and performance (Scale 11). Ratings of xerox and mail services were unrelated to any of the GPSA summary scale scores; this was most likely due to the fact these these services were generally rated as available and adequate by faculty.

All five of the criterion measures of faculty scholarship and productivity were related to one or both of the GPSA faculty productivity scales (Scales 15 and 16). The four measures of faculty publication history were positively correlated to Scale 15, faculty research activities. Number of presentations for the last two years was postitively related to both Scale 15 and Scale 16, faculty professional activities.

With the doctoral program as the unit of analysis (see Table 11), the four criterion measures are primarily indicators of faculty scholarship and productivity. As would be expected, these measures correlated very positively with mean faculty ratings of their program's scholarly excellence (Scale 2) and their research activities (Scale 15), demonstrating strong evidence for concurrent validity of these scales. When faculty ratings of their doctoral program were aggregated (averaged) to the program level, it is interesting to note that rankings by faculty of only nursing doctoral programs (1984 cooperative program evaluation) were statistically related to 6 of the 11 GPSA summary scales,

whereas rankings by deans and other nursing leaders of all nursing schools (Chamings, 1984) were not statistically related to any of the scales.

Student questionnaire. Descriptive statistics and correlations of criterion measures with GPSA summary scale scores for students are reported in Table 12 (individual respondent as the unit of analysis) and Table 13 (program as the unit of analysis). The set of academic and social environment measures for students was limited to the five environmental descriptors within the set of items developed by the national advisory committee. Three of these descriptors ("scholarly," "social," and "healthy") were positively correlated with scores on at least 6 of the 10 summary scales for students. Only Scale 7 (available resources) and Scale 8 (student committment and motivation) demonstrated no relationships with these five environmental descriptors. It is interesting to note that the "social" descriptor was positively related to many GPSA summary scale scores for students, but not for faculty; whereas, the "prestigious" descriptor was positively related to most scale scores for faculty, but not for students.

Insert Tables 12 and 13 about here

There were no criterion measures of program resources and management for students. Criterion measures of student scholarship and productivity included three measures of publication history, number of presentations for the last two years, and whether or not a student received an Advanced Nurse Traineeship or NRSA Pre-doctoral Fellowship; none of these measures correlated with any of the student GPSA summary scales. This is not surprising, however, for the 10 student summary scales are primarily indicators of the academic program environment as perceived by students; there are no student productivity scales, such as Scales 15 and 16 for faculty.

With the doctoral program as the unit of analysis (see Table 13), three of the four indicators of faculty scholarship and productivity correlated positively with mean student ratings of their program's scholarly excellence (Scale 2), as was found with faculty ratings. Several of these indicators also related positively to available resources (Scale 7) and student assistantship experiences (Scale 10), although only one correlation was statistically significant. One possible explanation for these associations is that Division of Nursing (DON) grant funding most likely provided greater availablility of resources for students, particularly salary support for research and teaching assistantships. This in turn may have increased both the quantity and quality of assistantship experiences for students

in these doctoral programs, as well as their overall satisfaction with the value of their educational experiences.

Alumni questionnaire. Descriptive statistics and correlations of criterion measures with GPSA summary scale scores for alumni are reported in Table 14 (individual respondent as the unit of analysis) and Table 15 (program as the unit of analysis). Within the set of academic and social environment measures, three of the five environmental descriptors ("scholarly," "healthy", and "prestigious") were positively correlated with scores on most of the 10 summary scales for alumni. As with students, Scale 7 (available resources) demonstrated no relationships with any of the five environmental descriptors. It is interesting to note that the three descriptors positively correlated with GPSA summary scales for alumni are the same as those for faculty; the "social" descriptor, correlated in student ratings, appears to be less associated with faculty and alumni ratings of the environment.

Insert Tables 14 and 15 about here

One item in the GPSA alumni questionnaire asks respondents to rate, overall, how well their department prepared them for their primary purpose in pursuing a doctoral degree. This global indicator of alumni satisfaction with the quality of their doctoral education correlated positively with all but one (Scale

7) of the GPSA summary scales. As would be expected, alumni who felt that they received better preparation rated the environment of their doctoral program more positively.

There were no criterion measures of program resources and management for alumni. Criterion measures of alumni scholarship and productivity included two measures of publication history, number of presentations for the last two years, and whether or not alumni received an Advanced Nurse Traineeship or NRSA Pre-doctoral Fellowship when students; none of these measures correlated with any of the alumni GPSA summary scales. Once again, this is not surprising, for the 10 alumni summary scales are primarily indicators of the academic program environment as perceived by alumni; there are no GPSA productivity scales for alumni (or students).

With the doctoral program as the unit of analysis (see Table 15), two of the four indicators of faculty scholarship and productivity correlated positively with mean alumni ratings of their program's scholarly excellence (Scale 2), as was found with both faculty and student ratings. Two of the indicators related positively to available resources (Scale 7), and one indicator (number of DON funded research grants) also related positively to value of educational experience for employment (Scale 14). These results were similar to those found with students and are discussed in the previous section of the paper.

Summary and Recommendations

Content validity of the questionnaires was substantiated by various groups of experts in nursing doctoral education. They found the GPSA instruments to have content appropriate and valid for evaluating a variety of dimensions of quality common to all doctoral programs. They noted, however, an important limitation of the questionnaires: items fail to assess areas of concern to practice disciplines, such as advanced clinical practice in nursing. This limitation is somewhat remedied by the GPSA option of allowing up to 10 locally developed, fixed-format items to be added to each of the questionnaires.

Factorial construct validity analyses at the item level indicated that most of the summary scales demonstrated scale homogeneity, that is, they were measuring one primary factor or construct. Only Scale 15 (faculty research activities) demonstrated factorial complexity, with three items defining a moderately strong first factor related to grant support of faculty research and three items defining a somewhat weaker second factor related to recognition of excellence in research and scholarly writing. Consideration should be given to splitting Scale 15 into two new scales of three items each, and then adding new items tapping the concepts of the two new scales.

Internal-consistency reliability analyses also indicated that most of the summary scales demonstrated satisfactory scale

homogeneity and, therefore, internal-consistency reliability for group-level data, based on coefficient alphas .60 or greater. Only Scale 7 (available resources) for alumni and Scale 16 (faculty professional activities) demonstrated a lack of internal consistency. Scale 7 contains only two items in the student and alumni questionnaires and three items in the faculty questionnaire. Clearly, the reliability of the scale could be improved for all three questionnaires by the addition of more items tapping the same concept.

The relatively low coefficent alpha of .49 for Scale 16 is not surprising given the low intercorrelations of all five items making up the scale. Both Scales 15 and 16 are reported in percentages of items in those scales to which faculty responded positively in the list of individual research and professional activities presented. This is in contrast to Scales 1 through 14, which are reported as averages of the Likert-scaled item responses making up those scales. Perhaps the validity and reliability problems concerning these two scales are related more to the type of item scaling (yes/no) and the validity and reliability analytic methods chosen, rather than to the content of the items making up the two scales. Further investigation into the content and structure of these two scales is warranted. It is clear that measures of productivity are important at the GPSA summary scale

level, and consideration should also be given to developing similar scales for the student and alumni questionnaires.

Based on the results of the item-level factorial validity and reliability analyses, several summary scales demonstrated higher coefficient alphas (and, therefore, increased reliability) when items that did not seem to relate to other items within the scale were excluded. These included Scale 1 for the faculty, student, and alumni questionnaires (drop I-3 and I-7); Scale 6 for faculty (drop I-9); Scale 11 for faculty (drop I-12); and Scale 12 for faculty (drop I-5). It must be noted that all recommendations for summary scale changes are based on the results of analyses with samples of nursing faculty, students, and alumni. The comprehensiveness of the samples (25 of 29 nursing doctoral programs in 1983-84 participating; 54% to 60% response rates for individual faculty, students, and alumni) lends support to the external validity of the findings for nursing doctoral education. Similar analyses on data from doctoral programs in other disciplines will be necessary to demonstrate the generalizability of the recommendations.

Results of the second-order factor analyses using the 16 summary scale scores indicated that Scales 1 through 14 are related and measure various aspects of the academic program environment, whereas Scales 15 and 16 (faculty only) measure aspects of faculty productivity. The wide range of summary scale

intercorrelations within each questionnaire, however, [illegible] the concept of the summary scales as separate, though sufficiently related measures of the multidimensional concept of quality in doctoral education. This finding was also supported by the results of the concurrent validity analyses. Finally, associations between internal and external criterion measures and the GPSA summary scales demonstrated moderate evidence for concurrent validity of the scales.

In general, the faculty, student, and alumni GPSA questionnaires demonstrated satisfactory validity and reliability for evaluation of nursing doctoral education. The investigation into the psychometric properties of the instruments, with primary emphasis on the 6 summary scales, provided supportive evidence that there are indeed multiple dimensions of quality in doctoral education, and that those dimensions can be measured with the GPSA questionnaires and demonstrated to correlate with other measures of quality. The results of the current study add to the information that is currently available concerning the validity and reliability of the GPSA questionnaires for self-study and review of doctoral degree programs, and provide additional appropriate comparison data for another discipline, nursing.

References

Allen, M. J., & Yen, W. M. (1979). Introduction to Measurement Theory. Monterey, CA: Brooks/Cole.

Barhyte, D., & Holzemer, W. L. (1981). Cooperative evaluation in nursing doctoral education. Proceedings of the 1981 Forum on Doctoral Education in Nursing (Appendix). Seattle, Washington: University of Washington.

Blau, P. M., & Margulies, R. Z. (1974, Winter). The reputations of American professional schools. Change, ١ ٦. 42-45.

Carmines, E. G., & Zeller, R. A. (1979). Reliability and validity assessment. Beverly Hills, CA: Sage.

Cattell, R. B. (1965). Factor analysis: An introduction to essentials. Biometrics, 21, 190-215, 405-435.

Chamings, P. A. (1984). Ranking the nursing schools. Nursing Outlook, 32 (5), 238-239.

Clark, M. J. (1983). Graduate Program Self-Assessment Service for Doctoral Programs: Handbook for Users. Princeton, NJ: Educational Testing Service.

Clark, M. J., Hartnett, R. T., & Baird, L. L. (1976). Assessing Dimensions of Quality in Doctoral Education: A Technical Report of a National Study in Three Fields. Princeton, NJ: Educational Tesing Service.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences (rev. ed.). New York: Academic Press.

Fagin, C. M. (1982). The quality of nursing journals as rated by deans of nursing schools. Heart and Lung, 11, 65-68.

Gourman, J. (1980). The Gourman Report: A Rating of Graduate and Professional Programs in American and International Universities. Los Angeles: National Educational Standards.

Grout, J. W. (1985). Criticism of 'Institutional sources of articles published in 13 nursing journals, 1978-82' [Letter to the editor]. Nursing Research, 34, 204.

Guilford, J. P. (1965). Fundamental statistics in psychology and education (4th ed.). New York: McGraw-Hill.

Hayter, J. (1984). Institutional sources of articles published in 13 nursing journals, 1978-1982. Nursing Research, 33 (6), 357-362.

Holzemer, W. L. (1978). Towards a cooperative venture in program evaluation. Proceedings of the 1978 Forum on Doctoral Education in Nursing (pp. 65-71). Chicago: Rush University.

Holzemer, W. L. (1982). Quality in graduate nursing education. Nursing and Health Care, 3 (11), 171-189.

Holzemer, W. L. (1987). Doctoral education in nursing: An assessment of quality, 1979-1984. Nursing Research, 36 (2), 111-116.

Holzemer, W. L., & Barhyte, D. Y. (1979). Cooperative program evaluation project in doctoral education in nursing. Proceedings of the 1979 Forum on Doctoral Education in

Nursing (pp. 51-64). San Francisco: The University of California, San Francisco.

Holzemer, W., Barhyte, D., & Clark, M. (1980). Cooperative program evaluation project in doctoral education in nursing. Proceedings of the 1980 Forum on Doctoral Education in Nursing (pp. 68-109). Detroit, Michigan: Wayne State University.

Holzemer, W. L., & Chambers, D. B. (1986). Healthy nursing doctoral programs: Relationship between perceptions of the academic environment and productivity of faculty and alumni. Research in Nursing & Health, 9 (4), 299-307.

Holzemer, W. L., & Chambers, D. B. (1988). A contextual analysis of faculty productivity. Journal of Nursing Education, 27 (1), 10-18.

Kaiser, H. F. (1974). An index of factorial simplicity. Psychometrika, 39, 31-36.

Pelczar, M. J. (1985). The nature of the doctorate and criteria for quality. Journal of Professional Nursing, 1 (2), 94-100.

Polit, D. F., & Hungler, B. P. (1978). Nursing research: Principles and methods. Philadelphia: Lippincott.

Thorndike, R. L., & Hagen, E. P. (1977). Measurement and evaluation in psychology and education (4th ed.). New York: John Wiley.

Table 1

Overall Response Rates for the 1984 GPSA Questionnaire

	Faculty	Student	Alumni
Number Distributed	592	1229	494
Number Returned	329	669	296
Number Usable [a]	326	659	296
Response Rate [b]	55%	54%	60%
Number of Applicable Programs	25	24	18
Number of Programs with a Minimum of 5 Returned and Usable Questionnaires [c]	25	22	16

[a] A usable questionnaire is defined as any GPSA questionnaire having valid responses to 10 or more questions across Parts I and II combined (criteria set by ETS)

[b] Response Rate = Number Usable / Number Distributed (criteria set by ETS)

[c] Criteria set by ETS for calculating program mean scores for inclusion in the ETS GPSA Program Report

Table 2

Description of 16 GPSA Summary Scales

1. Environment for Learning. The extent to which the department provides a supportive environment characterized by mutual respect and concern between students and professors, students' helpfulness to one another, and department openness to new ideas and different points of view. (6 items)
2. Scholarly Excellence. Rated excellence of the department faculty, ability of students, and intellectual stimulation in the program. (5 items)
3. Quality of Teaching. Faculty excitement for new ideas and helpfulness in dealing with class work; student evaluation of faculty teaching methods, grading procedures, and preparation for class. (7 items)
4. Faculty Concern for Students. The extent to which faculty members are perceived to be interested in the welfare and professional development of students, accessible, and aware of student needs, concerns, and suggestions. (5 items)
5. Curriculum. Ratings of the variety and depth of graduate course and program offerings, program flexibility, opportunities for individual projects, and interactions with related departments. (5 items)

(table continues)

6. Departmental Procedures. Ratings of departmental policies and procedures such as the relevance and administration of degree requirements, evaluation of student progress toward the degree, academic advisement of students, and helpfulness to graduates in finding appropriate employment. (8 items faculty, 10 items students, 9 items alumni)

7. Available Resources. Ratings of available facilities such as libraries and laboratories, and overall adequacy of physical and financial resources for a doctoral program. (3 items faculty, 2 items students and alumni)

8. Student Commitment and Motivation. Judgments about the extent to which doctoral students do a lot of unassigned reading, demonstrate enthusiastic involvement with the field, carefully prepare for courses, and persist on projects despite setbacks. (4 items)

9. Student Satisfaction with Program. Self-reported student satisfaction with the program as reflected in judgments about the amount that has been learned, preparation for intended career, desire to transfer, and willingness to recommend the program to a friend. (4 items students, 3 items alumni)

10. Student Assistantship or Internship Experiences. Ratings of preparation for and supervision of assigned duties; contribution of the experiences to academic and professional development. (7 items)

(table continues)

11. Departmental Direction and Performance. Faculty judgments about teaching practices in the department, and about departmental management in areas such as the career development of junior faculty, planning, and administration. (7 items)

12. Faculty Work Environment. Self-reported faculty satisfaction with departmental objectives and procedures, academic freedom, opportunities to influence decisions, and relationships with other faculty members; sense of conflicting demands and personal strain. (6 items)

13. Alumni Dissertation Experiences. Judgments about the ways in which dissertation topics were identified and committees appointed, interactions with the committee, standards of performance, and relationship of the experience to other professional skills and employment demands. (11 items)

14. Value of Educational Experiences for Employment. Alumni judgments about their graduate school experiences as preparation for present work demands in areas such as required and elective courses, associations with faculty members and students, departmental standards, and gains in specific knowledge or skills. (13 items)

(table continues)

15. Faculty Research Activities. The extent to which faculty members report receiving awards for outstanding research or scholarly writing, editing professional journals, refereeing articles submitted to professional journals, and receiving grants to support research or other scholarly or creative work. (6 items)

16. Faculty Professional Activities. The extent to which faculty members report serving on national review or advisory councils, holding office in regional or national professional associations, and receiving awards for outstanding teaching or professional practice. (5 items)

Table 3

GPSA Faculty Questionnaire

Descriptive Statistics and Intercorrelations for Summary Scale Scores

GPSA Summary Scales	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Environment for Learning	3.26	0.51	100															
2. Scholarly Excellence	3.31	0.62	62	100														
3. Quality of Teaching	---	---	---															
4. Faculty Concern for Students	3.24	0.52	73	58	--	100												
5. Curriculum	3.20	0.58	63	61	--	57	100											
6. Departmental Procedures	3.25	0.51	73	70	--	67	71	100										
7. Available Resources	2.88	0.72	40	53	--	35	48	53	100									
8. Student Commitment and Motivation	3.46	0.53	59	68	--	61	55	62	45	100								
9. Student Satisfaction with Program	---	---	---															
10. Student Assistantship Experiences	---	---	---															
11. Departmental Direction and Performance	3.07	0.54	73	76	--	66	69	81	51	60	--	--	100					
12. Faculty Work Environment	3.09	0.61	72	64	--	53	57	62	42	49	--	--	69	100				
13. Alumni Dissertation Experiences	---	---	---															
14. Value of Educa. Exper. for Employment	---	---	---															
15. Faculty Research Activities	51%	29%	04	16	--	03	12	09	18	10	--	--	12	10	--	--	100	
16. Faculty Professional Activities	50%	28%	14	17	--	07	22	19	14	16	--	--	20	16	--	--	30	100

(table continues)

Note. Descriptive statistics and intercorrelations are based on the 299 faculty who had ETS-calculated scale scores for all 11 faculty GPSA summary scales. Dashed lines indicate scales not applicable to the faculty questionnaire. Decimal points for correlation coefficients not printed.

Table 4

GPSA Faculty Questionnaire

Internal Consistency Reliability and Factorial Validity Analyses for Summary Scale Scores

GPSA Summary Scales (# Items)	Interitem Correlations			Coefficient Alpha	Factors Retained			
					Factor 1		Factor 2	
	Min.	Max.	Mean		% Var.[a]	Item Load.[b]	% Var.[a]	Item Load.[b]
1. Environ. for Learning (6)	-08	55	31	73	45	31 to 77	----------	
2. Scholarly Excellence (5)	51	72	61	89	69	70 to 82	----------	
4. Fac. Concern for Students (5)	30	58	45	73	56	54 to 81	----------	
5. Curriculum (5)	33	73	44	79	56	54 to 73	----------	
6. Departmental Procedures (8)	18	63	44	86	51	38 to 75	----------	
7. Available Resources (3)	28	48	41	68	61	52 to 88	----------	
8. Student Commit./Motiva. (4)	48	73	55	81	66	63 to 85	----------	
11. Depart. Direct./Perform. (7)	20	63	41	82	50	42 to 74	----------	
12. Faculty Work Environment (6)	15	58	38	77	50	27 to 84	----------	
15. Fac. Research Activities (6)	06	53	21	62	35	08 to 74	20	05 to 77
16. Fac. Professional Activ. (5)	06	28	16	49	33	26 to 60	----------	

(table continues)

Note. The reliability and factorial validity analyses for all scales are based on the 236 faculty who answered all 60 summary scale items of the GPSA Faculty Questionnaire. Decimal points not printed.

[a] Before rotation

[b] After varimax rotation if more than one extracted factor retained from initial solution

Table 5

GPSA Student Questionnaire

Descriptive Statistics and Intercorrelations for Summary Scale Scores

GPSA Summary Scales	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Environment for Learning	3.05	0.51	100															
2. Scholarly Excellence	3.34	0.57	66	100														
3. Quality of Teaching	3.01	0 61	67	78	100													
4. Faculty Concern for Students	2.91	0.69	76	69	74	100												
5. Curriculum	2.99	0.62	62	71	75	70	100											
6. Departmental Procedures	2.99	0.57	67	71	77	74	81	100										
7. Available Resources	3.00	0.79	29	40	38	30	45	45	100									
8. Student Commitment and Motivation	3.54	0.43	43	47	45	38	40	45	20	100								
9. Student Satisfaction with Program	3.47	0.61	64	77	70	66	69	70	37	37	100							
10. Student Assistantship Experiences	2.94	0.72	51	51	51	55	47	57	34	23	44	100						
11. Departmental Direction and Performance	----		----															
12. Faculty Work Environment	----		----															
13. Alumni Dissertation Experiences	----		----															
14. Value of Educa. Exper. for Employment	----		----															
15. Faculty Research Activities	----		----															
16. Faculty Professional Activities	----		----															

(table continues)

Note. Descriptive statistics and intercorrelations among Scales 1 through 9 are based on the 538 students who had ETS-calculated scale scores for all 9 of these student GPSA summary scales. Descriptive statistics and intercorrelations of Scale 10 with Scales 1 through 9 are based on the 252 students who had been a research or teaching assistant in their department and had ETS-calculated scale scores for all 10 of the student GPSA summary scales. Dashed lines indicate scales not applicable to the student questionnaire. Decimal points for correlation coefficients not printed.

Table 6

GPSA Student Questionnaire

Internal Consistency Reliability and Factorial Validity Analyses for Summary Scale Scores

GPSA Summary Scales (# Items)	Interitem Correlations			Coefficient Alpha	Factors Retained			
					Factor 1		Factor 2	
	Min.	Max.	Mean		% Var.[a]	Item Load.[b]	% Var.[a]	Item Load.[b]
1. Environ. for Learning (6)	01	63	27	69	42	14 to 80	----------	
2. Scholarly Excellence (5)	38	73	56	86	66	54 to 89	----------	
3. Quality of Teaching (7)	45	72	58	91	64	67 to 83	----------	
4. Fac. Concern for Students (5)	50	74	62	89	70	73 to 86	----------	
5. Curriculum (5)	32	75	50	84	61	53 to 78	----------	
6. Departmental Procedures (10)	14	74	44	89	51	48 to 84	----------	
7. Available Resources (2)	49	49	49	66	75	70 (2 items)	----------	
8. Student Commit./Motiva. (4)	24	57	36	67	53	36 to 82	----------	
9. Student Satis. with Prog. (4)	53	70	63	87	73	74 to 89	----------	
10. Student Assistant. Exper. (7)	27	74	46	86	55	47 to 80	----------	

(table continues)

Note. The reliability and factorial validity analyses for Scales 1 through 9 are based on the 293 students who answered all 48 items comprising Scales 1 through 9 of the GPSA Student Questionnaire. The analyses for Scale 10 are based on the 281 students who had been a research or teaching assistant in their department and answered all 7 items comprising Scale 10. Decimal points not printed.

[a] Before rotation

[b] After varimax rotation if more than one extracted factor retained from initial solution

Table 7

GPSA Alumni Questionnaire

Descriptive Statistics and Intercorrelations for Summary Scale Scores

GPSA Summary Scales	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Environment for Learning	3.17	0.51	100															
2. Scholarly Excellence	3.34	0.59	52	100														
3. Quality of Teaching	3.12	0.61	63	75	100													
4. Faculty Concern for Students	3.07	0.69	74	60	70	100												
5. Curriculum	3.18	0.60	53	60	72	62	100											
6. Departmental Procedures	3.11	0.58	61	64	80	67	72	100										
7. Available Resources	3.20	0.73	20	36	35	24	46	38	100									
8. Student Commitment and Motivation	----------		----------															
9. Student Satisfaction with Program	3.52	0.59	50	75	69	53	59	64	27	--	100							
10. Student Assistantship Experiences	----------		----------															
11. Departmental Direction and Performance	----------		----------															
12. Faculty Work Environment	----------		----------															
13. Alumni Dissertation Experiences	3.32	0.53	46	59	67	56	64	76	27	--	58	--	--	--	100			
14. Value of Educa. Exper. for Employment	3.26	0.49	39	64	61	48	59	65	40	--	61	--	--	--	66	100		
15. Faculty Research Activities	----------		----------															
16. Faculty Professional Activities	----------		----------															

(table continues)

Note. Descriptive statistics and intercorrelations are based on the 260 alumni who had ETS-calculated scale scores for all 10 alumni GPSA summary scales. Dashed lines indicate scales not applicable to the alumni questionnaire. Decimal points for correlation coefficients not printed.

Table 8

GPSA Alumni Questionnaire

Internal Consistency Reliability and Factorial Validity Analyses for Summary Scale Scores

GPSA Summary Scales (# Items)	Interitem Correlations			Coefficient Alpha	Factors Retained			
					Factor 1		Factor 2	
	Min.	Max.	Mean		% Var.[a]	Item Load.[b]	% Var.[a]	Item Load.[b]
1. Environ. for Learning (6)	07	57	29	71	43	24 to 77	----------	
2. Scholarly Excellence (5)	47	70	57	87	66	63 to 84	----------	
3. Quality of Teaching (7)	36	68	53	88	60	65 to 83	----------	
4. Fac. Concern for Students (5)	40	78	58	86	67	61 to 91	----------	
5. Curriculum (5)	31	67	44	79	55	58 to 77	----------	
6. Departmental Procedures (9)	21	67	44	87	51	48 to 79	----------	
7. Available Resources (2)	40	40	40	57	70	63 (2 items)	----------	
9. Student Satis. with Prog. (3)	32	64	49	73	67	no solution	----------	
13. Alumni Disserta. Exper. (11)	30	72	44	89	49	57 to 77	----------	
14. Value of Educa. for Employ. (13)	04	77	32	85	38	39 to 77	----------	

(table continues)

Note. The reliability and factorial validity analyses for Scales 1 through 13 are based on the 207 alumni who answered all 54 items comprising Scales 1 through 13 of the GPSA Alumni Questionnaire. Because of scoring rules for Scale 14 set by Educational Testing Service (ETS), the analyses for Scale 14 are based on the 68 alumni who had been a research or teaching assistant in their department and answered all 13 items comprising Scale 14. Decimal points not printed.

[a] Before rotation

[b] After varimax rotation if more than one extracted factor retained from initial solution

Table 9

Descriptive Statistics and Intercorrelations for the 3 External Criterion Measures and the Faculty Ranking of Doctoral Programs (Internal Measure)

Criterion Measures	Mean	Range	N	Intercorrelations 1	2	3	4
1. Ranking of doctoral programs in nursing by faculty (1984 cooperative program evaluation)	13.0	1-25	25	100			
2. Ranking of all nursing schools by deans and nursing academics and professionals (Chamings, 1984)	14.0	1-32	22	84	100		
3. Number of faculty publications in scholarly nursing journals, 1978-1982 (Grout, 1985)	7.4	0-29	25	55	54	100	
4. Number of Division of Nursing (DON) funded research grants, 1979-1983	2.9	0-12	24	56	48	72	100

Note. Unit of analysis is the program. All Spearman rank-order correlations were significant at $p \leq .05$, two-tailed. Decimal points for correlation coefficients not printed.

Table 10

GPSA Faculty Questionnaire

Concurrent Validity Analysis: Descriptive Statistics and Correlations of Criterion Measures with Summary Scale Scores using the Individual Respondent as the Unit of Analysis

Criterion Measures	Mean	SD	N	Correlations with Faculty GPSA Summary Scales [a]										
				1	2	4	5	6	7	8	11	12	15	16
Academic and Social Environment														
Academic rank (1=no rank, 6=full professor)	5.1	0.8	319											38
Tenure (1=no, 2=yes)	1.6	0.5	320											37
Described environment of doctoral program as: (1=not at all, 4=extremely)														
Stressful	2.4	0.8	270	-32		-31						-34		
Scholarly	3.0	0.8	270	38	58	38	36	41	34	41	53	36		
Social	2.0	0.7	271			35								
Healthy	2.4	0.7	285	46	35	41	38	42		30	46	50		
Prestigious	2.8	1.0	282	31	55	34		40	37	33	45	33		
% time teaching/advising students	45.6	22.7	320											
% time research/scholarly work	24.9	15.2	320										41	
% time admininstration/consulting/other	29.6	24.0	320											
% of colleag. in active program of research	64.6	30.5	308	34	51	34	35	40		32	37	33		

(table continues)

Criterion Measures	Mean	SD	N	Correlations with Faculty GPSA Summary Scales [a] 1	2	4	5	6	7	8	11	12	15	16
Resources and Management														
Rated adequacy of following support services:														
(0=not available, 3=excellent)														
Xerox	2.3	0.7	313											
Mailroom services	2.2	0.6	308											
Secretarial support	1.9	0.7	311									32		
Travel monies	1.2	0.8	311									30		
Express mail services	2.0	0.9	274											
Release time for scholarly activity	1.5	0.9	298				38				30	30		
Scholarship and Productivity														
Total publications for entire career	19.8	28.1	315										35	
Total publications for last 3 years	7.5	7.7	315										46	
# refereed articles published entire career	8.8	10.2	286										35	
# refereed articles published last 3 years	3.7	3.9	288										38	
Total presentations last 2 years	8.1	9.5	308										42	32

(table continues)

Note. Only Pearson product-moment correlations with an absolute value of .30 or greater are reported. Because of the large sample sizes, all correlations of this magnitude were significant at $p <= .001$, two-tailed. Decimal points for correlation coefficients not printed.

[a] See Table 2 for description of the 16 GPSA Summary Scales.

Table 11

GPSA Faculty Questionnaire

Concurrent Validity Analysis: Correlations of Criterion Measures with Summary Scale Scores using the Program as the Unit of Analysis

Criterion Measures	Correlations with Faculty GPSA Summary Scales [a]										
	1	2	4	5	6	7	8	11	12	15	16
Ranking of doctoral programs in nursing by faculty (1984 cooperative program evaluation)		71	33	43	30	48	57	47		42	
Ranking of all nursing schools by deans and nursing academics and professionals (Chamings, 1984)		33						30			
Number of faculty publications in scholarly nursing journals, 1978-1982 (Grout, 1985)		55					35			61	
Number of Division of Nursing (DON) funded research grants, 1979-1983		51					34			63	

Note. With the program as the unit of analysis, sample sizes for the correlations varied from 22 to 25. Only Spearman rank-order correlations with an absolute value of .30 or greater are reported; correlations significant at $p <= .05$, two-tailed, are underlined. Decimal points for correlation coefficients not printed.

[a] See Table 2 for description of the 16 GPSA Summary Scales.

Table 12

GPSA Student Questionnaire

Concurrent Validity Analysis: Descriptive Statistics and Correlations of Criterion Measures with Summary Scale Scores using the Individual Respondent as the Unit of Analysis

Criterion Measures	Mean	SD	N	Correlations with Student GPSA Summary Scales [a]									
				1	2	3	4	5	6	7	8	9	10
Academic and Social Environment													
Described environment of doctoral program as:													
(1=not at all, 4=extremely)													
Stressful	2.7	0.8	530	43			38	31	32				
Scholarly	3.2	0.8	518		53	46	35	40	34			44	
Social	2.0	0.7	473	31	32		36	32	34			31	
Healthy	2.4	0.8	495	53	39	50	55	51	52			45	33
Prestigious	3.0	0.9	521		37								
Scholarship and Productivity													
Total publications	2.5	4.4	647										
# refereed articles published entire career	1.3	3.2	599										
# refereed articles published last 3 years	0.8	1.3	581										
Total presentations last 2 years	2 3	3.6	561										

(table continues)

Criterion Measures	Mean	SD	N	Correlations with Student GPSA Summary Scales [a] 1	2	3	4	5	6	7	8	9	10
Scholarship and Productivity (cont.)													
Received financial aid in form of:													
(0=no, 1=yes)													
Advanced Nurse Traineeship	0.4	0.5	668										
NRSA Pre-doctoral Fellowship	0.1		668										

Note. Only Pearson product-moment correlations with an absolute value of .30 or greater are reported. Because of the large sample sizes, all correlations of this magnitude were significant at $p <= .001$, two-tailed. Decimal points for correlation coefficients not printed.

[a] See Table 2 for description of the 16 GPSA Summary Scales.

Table 13

GPSA Student Questionnaire

Concurrent Validity Analysis: Correlations of Criterion Measures with Summary Scale Scores using the Program as the Unit of Analysis

Criterion Measures	Correlations with Student GPSA Summary Scales [a]									
	1	2	3	4	5	6	7	8	9	10
Ranking of doctoral programs in nursing by faculty (1984 cooperative program evaluation)		56					43			31
Ranking of all nursing schools by deans and nursing academics and professionals (Chamings, 1984)										
Number of faculty publications in scholarly nursing journals, 1978-1982 (Grout, 1985)		47								39
Number of Division of Nursing (DON) funded research grants, 1979-1983		55					40			34

Note. With the program as the unit of analysis, sample sizes for the correlations varied from 19 to 22. Only Spearman rank-order correlations with an absolute value of .30 or greater are reported; correlations significant at $p \leq .05$, two-tailed, are underlined. Decimal points for correlation coefficients not printed.

[a] See Table 2 for description of the 16 GPSA Summary Scales.

Table 14

GPSA Alumni Questionnaire

Concurrent Validity Analysis: Descriptive Statistics and Correlations of Criterion Measures with Summary Scale Scores using the Individual Respondent as the Unit of Analysis

Criterion Measures	Mean	SD	N	Correlations with Alumni GPSA Summary Scales [a]									
				1	2	3	4	5	6	7	9	13	14
Academic and Social Environment													
Described environment of doctoral program as:													
(1=not at all, 4=extremely)													
Stressful	2.6	0.8	264										
Scholarly	3.3	0.8	259	34	62	49	36	43	46		47	50	52
Social	2.2	0.8	247				47	31					35
Healthy	2.4	0.8	257	48	34	40	47	32	43		33	39	34
Prestigious	3.0	1.0	263		49	40		36	40		32	39	41
Overall, how well department prepared for primary purpose in pursuing degree (1=not very well, 3=extremely well)	2.6	0.6	295	32	57	55	35	47	55		58	57	58
Scholarship and Productivity													
Total publications for entire career	9.0	20.6	274										
Total publications for last 3 years	4.5	5.5	274										
Total presentations last 2 years	7.3	8.5	287										

(table continues)

Criterion Measures	Mean	SD	N	Correlations with Student GPSA Summary Scales [a] 1	2	3	4	5	6	7	8	9	10
Scholarship and Productivity (cont.)													
Received financial aid in form of:													
(0=no, 1=yes)													
Advanced Nurse Traineeship	0.5	0.5	299										
NRSA Pre-doctoral Fellowship	0.1	0.3	299										

Note. Only Pearson product-moment correlations with an absolute value of .30 or greater are reported. Because of the large sample sizes, all correlations of this magnitude were significant at $p <= .001$, two-tailed. Decimal points for correlation coefficients not printed.

[a] See Table 2 for description of the 16 GPSA Summary Scales.

Table 15

GPSA Alumni Questionnaire

Concurrent Validity Analyses: Correlations of Criterion Measures with Summary Scale Scores using the Program as the Unit of Analysis

Criterion Measures	Correlations with Alumni GPSA Summary Scales [a]									
	1	2	3	4	5	6	7	9	13	14
Ranking of doctoral programs in nursing by faculty (1984 cooperative program evaluation)		44			35		52			
Ranking of all nursing schools by deans and nursing academics and professionals (Chamings, 1984)										
Number of faculty publications in scholarly nursing journals, 1978-1982 (Grout, 1985)		58			38				35	30
Number of Division of Nursing (DON) funded research grants, 1979-1983		51			36	31	62		45	51

Note. With the program as the unit of analysis, sample sizes for the correlations varied from 15 to 16. Only Spearman rank-order correlations with an absolute value of .30 or greater are reported; correlations significant at $p <= .05$, two-tailed, are underlined. Decimal points for correlation coefficients not printed.

[a] See Table 2 for description of the 16 GPSA Summary Scales.